Trends in Ecology & Evolution

CellPress REVIEWS

Opinion

Detecting the True Extent of Introgression during Anthropogenic Hybridization

S. Eryn McFarlane [1,2,*] and Josephine M. Pemberton[1]

Hybridization among naturally separate taxa is increasing owing to human impact, and can result in taxon loss. Previous classification of anthropogenic hybridization has largely ignored the case of bimodal hybrid zones, in which hybrids commonly mate with parental species, resulting in many backcrossed individuals with a small proportion of introgressed genome. Genetic markers can be used to detect such hybrids, but until recently too few markers have been used to detect the true extent of introgression. Recent studies of wolves and trout have employed thousands of markers to reveal previously undetectable backcrosses. This improved resolution will lead to increased detection of late-generation backcrosses, shed light on the consequences of anthropogenic hybridization, and pose new management issues for conservation scientists.

Highlights

Anthropogenic hybridization is increasingly common and is likely to result in a breakdown of reproductive isolation between 'good' species.

Backcrossed individuals that have only a small proportion of one parental genome are difficult to differentiate from parental individuals by using the most common current technologies.

Bimodal hybrid zones are characterized by introgression and backcrossing. The majority of hybrid individuals in these systems have low levels of introgression. The problems posed by bimodal hybrid zones have been largely overlooked in the literature.

Genome-wide sampling of genetic markers at high densities allow increased precision in the estimate of admixture proportions, which makes it feasible to detect multi-generation backcrosses, and will thus make it easier to differentiate bimodal hybrid zones from hybrid swarms or systems without introgression.

Anthropogenic Hybridization

Anthropogenic hybridization (see Glossary), in which human disturbance leads to range overlap and **hybridization** of previously reproductively isolated populations or species, is a growing conservation concern [1–3]. With increased human-generated movement of species into new ranges, there is an increasing number of cases of hybridization between species that were historically **allopatric** [4]. Disturbance of habitats can also result in a breakdown of reproductive isolation between previously isolated, **sympatric** species [1]. **Introgression** is usually difficult to detect from phenotypes, and there is growing evidence that backcrossing has often proceeded further than is detectable by low-density genetic marker panels. In this article we make the case that genomic approaches are essential and increasingly available to disentangle late-generation backcrosses from parental populations after introgression has occurred.

The Benefits of Anthropogenic Hybridization

There are possible benefits of anthropogenic hybridization. Policy-makers can use hybridization as a management tool to help endangered populations. In 'genetic rescue' programs (i.e., breeding programs designed to release small populations from inbreeding depression), individuals from a closely related population or subspecies are introduced to an inbred population to manage inbreeding depression. For example, when Florida panthers (*Puma concolor coryi*) were threatened due to inbreeding depression, eight Texas panthers (*P. concolor cougaur*) were introduced. The **hybrid** kittens survived better, and the population is now recovering [5]. Approximately 90% of such genetic rescue attempts have been successful, showing that anthropogenic hybridization is a viable conservation method [6]. Adaptive introgression ('evolutionary rescue'), in which beneficial alleles from an introduced population are selected for in hybrid individuals, is another possible benefit of anthropogenic hybridization. For example, a segment of chromosome 15 that has naturally introgressed from *Populus balsamifera* into *P. trichocarpa* appears to allow *P. trichocarpa* to live in colder, drier areas than *P. trichocarpa*

[1]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK
[2]Department of Biology, Lund University, Sölvegatan 37, 223 62 Lund, Sweden

*Correspondence: eryn.mcfarlane@ed.ac.uk (S.E. McFarlane).

individuals lacking this haplotype [7]. This suggests that there is potential for adaptive introgression to facilitate evolutionary rescue of populations at risk of extinction due to climate change [8], although such genomic management of at-risk populations will require much enabling research, and should be approached with caution [9,10].

The Problems with Anthropogenic Hybridization

Anthropogenic hybridization can cause problems for native species. When no offspring or sterile offspring are produced, reproductive effort is wasted [11]. When fertile F1s are formed, introgression between the two previously diverged species is possible. There are two reasons why even low levels of introgression of non-native alleles are of concern from a conservation perspective. First, if all individuals of a species are hybrids then the species as it was is extinct. This has been termed 'extinction by hybridization' [11–15]. Note, however, that many copies of the native alleles may still be present in the population, provided that the population itself is large enough, and from a 'gene viewpoint' we may be content with this mode of conservation [16].

The second problem with hybridization is that introgression and recombination break up linked gene complexes, and non-native alleles that are favored (or are no longer in linkage with deleterious alleles) can be swept to fixation [17]. Although this leads to an initial increase in biodiversity (because alleles from both the native and non-native populations are present) native alleles are lost as non-native alleles sweep to fixation. If we again take a gene viewpoint of biodiversity, any alleles lost from the native population are a loss in biodiversity from the system. For example, non-native alleles at three of 68 genetic markers have gone to fixation in some populations of California tiger salamanders (*Ambystoma californiense*) after hybridization with barred tiger salamanders (*A. mavortium*) [18]. This has occurred in California tiger salamander populations that are nearly 100 km from the original barred tiger salamander introduction site, suggesting that these alleles have higher fitness than the native, California tiger salamander alleles that they have replaced [18].

Goals of Studies of Anthropogenic Hybridization

Studies of anthropogenic hybridization have different goals. A researcher might be interested to know if hybridization has occurred at all in a population to determine whether it should provide the breeding stock for new populations, and/or whether it should be quarantined because of hybridization. Relatively few informative markers are necessary to detect individuals of hybrid origin in any particular population because the detection of any non-native allele is a clear indication of hybridization [19].

However, if a researcher wishes to understand more about the underlying processes of hybridization and introgression, then many more markers are required. Specific goals might include selection of individuals for breeding programs, understanding the relationship between genotype and phenotype, understanding the type of hybrid system involved (see next section), and investigating mating patterns and fitness. For any of these goals, it is ideal to quantify individual **admixture** accurately, and to achieve this hundreds or thousands of informative markers may be required (see below).

Classifying Hybridization

To assist researchers and policy-makers in addressing anthropogenic hybridization, Allendorf and colleagues [11] categorized hybridization outcomes. Types 1–3 applied to naturally occurring hybridization, whereas types 4–6 applied to anthropogenic hybridization. Type 4 results in few or sterile F1 hybrids, and is characterized by wasted reproductive effort. Type 5 results in a **hybrid swarm** with widespread introgression into particular populations, but

Glossary

Accuracy: the proportion of identified hybrids that are actually of hybrid ancestry [33]. A low accuracy suggests a high rate of type I errors, in which parental species individuals are erroneously assigned as hybrids.
Admixture: the mixing of genomes from structured or diverged populations.
Allopatry: species in non-overlapping ranges.
Ancestry-informative markers: genetic markers with substantial allele-frequency differences across populations, which can be used to assign individuals to each population [63].
Anthropogenic hybridization: the breakdown of reproductive isolation between two species as a result of human action, including but not limited to species introduction, habitat disturbance, or escape of domestic species.
Bimodal hybrid zone: a hybridizing population in which preference for parental phenotypes, or scarcity of hybrids with which to mate, results in a population that includes few F1 hybrids, and many backcrossed individuals with a low level of introgression that often resemble the parental species in phenotype. Can be unimodal (if backcrossing is into only one parental species) or bimodal (backcrossing into both parental species) [22].
Credible interval: the range of possible values surrounding a point estimate, representing the uncertainty in the estimate.
Diagnostic markers: markers with fixed allele differences across populations.
d_{xy}: an absolute measure of genetic differentiation, calculated as the proportion of nucleotides that differ between two homologous sequences within the same or different population.
Efficiency: the proportion of correctly identified individuals in each group [33]. If the null hypothesis is that an individual is from the parental species rather than a hybrid individual, then low efficiency suggests a high rate of type II errors, in which hybrid individuals are incorrectly assigned as parental species.

some populations do not experience hybridization at all. Finally, type 6 results in a complete hybrid swarm following breakdown of reproductive isolation between species across all populations [11].

Three axes of variation determine the outcome of anthropogenic hybridization: differences in hybrid fitness, time since **secondary contact**, and the mating patterns of hybrids. Time since secondary contact and the mating patterns of hybrids were not explicitly considered in the original categorization of Allendorf *et al*. Type 4 differs from types 5 and 6 along an axis of hybrid fitness, where intrinsic postzygotic isolation affects hybrids in type 4, but not in types 5 or 6. This results in little to no backcrossing in type 4 hybrid zones because hybrids are extremely unfit compared to parental species. This decrease in hybrid fitness must be extreme because, even with a 90% decrease in fitness, the proportion of hybrids in a hybridizing population is expected to increase [20].

We suggest that the only difference between the types 5 and 6 of Allendorf *et al*. [11] is time since secondary contact. When an F1 reproduces, all its offspring and descendants are admixed to some extent [20]. If type 5 characterizes a system where only one or few populations have introgression, type 6 is the logical outcome of this same system, assuming random mating and sufficient time for migration between populations. Thus, we consider types 5 and 6 to be the same, both hybrid swarms with a breakdown of assortative mating, and in which hybrids have the same mating success as individuals of either of the parental species, and they are sufficiently common that hybrid $\times$ hybrid matings occur.

When there is a preference among hybrids for parental species phenotypes, or hybrids are very rare, we expect a different pattern of introgression. Backcrossing into the parental species leads to an increasingly large number of individuals with a small proportion ($<$10%) of their genome that is from the opposite species. As backcrossing continues, morphological differences between parental species and backcrossed individuals lessen, making it more and more difficult to detect a backcross using only phenotypic traits. This results in many hybrid individuals with very small proportions of another genome, although with a maintained bimodal distribution of trait values between the two parental species (Figure 1). From a conservation perspective, we consider this to be a worst-case scenario because these introgressed individuals are very difficult to detect. This can be contrasted with a general lack of assortative mating, in which hybrid individuals are as likely to breed with other hybrid individuals as with parental species (leading to a hybrid swarm), or, in the unlikely event of true assortative mating, where hybrid individuals preferentially breed with each other, which would lead to the eventual formation of a hybrid species (e.g., [21]). The contrast between hybrid zones with unimodal distributions of traits and admixture scores and those with bimodal distributions has previously been described in the context of naturally occurring hybrid zones [22], but does not yet seem to inform studies of anthropogenic hybridization.

The distribution of hybrid scores in a system at equilibrium varies depending on ecological factors that can affect hybrid fitness and hybrid encounter rate. Extrinsic postzygotic isolation can vary according to ecological factors that affect the ability of hybrids to successfully mate and reproduce [23]. Further, stochastic factors, particularly when hybrids are rare, and/or management might alter the reproductive success of hybrid individuals in wild systems. However, if hybrids are fertile, the proportion of hybrid individuals in all populations should increase [20], leading to the extreme endpoints of majority hybrid populations which either follow a hybrid swarm or **bimodal hybrid zone** distribution.

F_{st}: a measure of genetic differentiation between populations based on the difference in allele frequencies within and between populations [64].
Hybrid: an individual that has an intermediate genotype between two diverged, parental populations as the result of interbreeding between these populations.
Hybrid swarm: a hybridizing population that includes F1 hybrids and various backcrosses owing to a total breakdown of assortative mating. Also known as a unimodal hybrid zone [22].
Hybridization: mating of individuals from diverged populations.
Introgression: the movement of alleles between genetically differentiated forms (including populations, species, etc.) that is mediated by backcrossing [65].
Secondary contact: occurs when two (or more) species that have been in allopatry come back into sympatry.
Sympatry: species in overlapping ranges.

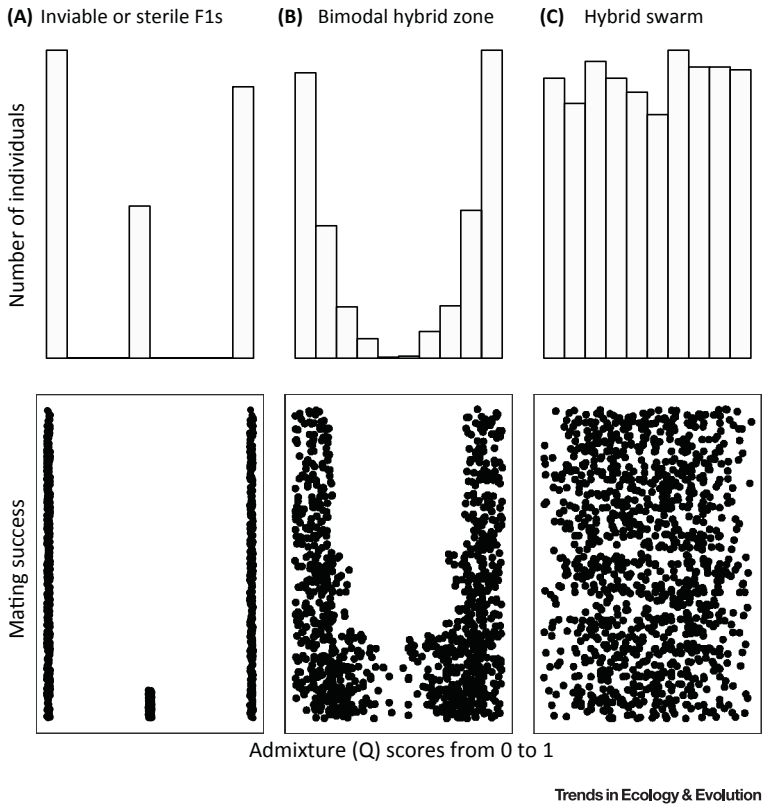

Figure 1. Anthropogenic Hybridization Falls into Three Main Categories. These are (A) systems with inviable or infertile hybrids, (B) bimodal hybrid zones in which there is either mating preference for parental species phenotypes or the relative abundance of parental species means most matings are backcrosses, and (C) hybrid swarms in which there is random mating and many hybrid individuals. In this schematic figure we illustrate for each type of anthropogenic hybridization system how many individuals of each admixture (*Q*) score might be found, and typical distributions of mating success across *Q* scores according to whether there is a high likelihood of hybrid individuals mating with the parental species that are present. Although we represent hybrid swarms and bimodal hybrid zones as being categorically different, these are probably the ends of a continuum and some systems may be intermediate between them. Note that we have represented (B) as a bimodal hybrid zone produced by backcrossing into both parental species. Alternatively there can be a single (i.e., unimodal) hybrid zone as a result of unidirectional backcrossing.

Key Considerations for Genetic Analyses of Anthropogenic Hybridization

Published studies of anthropogenic hybridization generally follow a similar protocol. Researchers use codominant marker genotypes to estimate divergence between the two species [24], and then use a clustering approach such as STRUCTURE [25–28] or ADMIXTURE [29,30] to partition individuals into different genetic groups (K). Those individuals with an admixture score (*Q*) intermediate between the extreme admixture scores associated with individuals of the parental species are designated hybrids. Many studies then use HYBRIDLAB [31,32] or similar methods to simulate hybrid genotypes from the sampled genotypes to assess the **efficiency** (i.e., type II error rate, rate of assigning hybrid individuals as parental species), and **accuracy** (i.e., type I error rate, rate of erroneously assigning parental species individuals as hybrids [33]). The 'overall performance' of an analysis is the product of efficiency and accuracy, and this performance can be used to assess the reliability of the study itself [33]. We outline here some best practices and points to consider to avoid underestimating the extent of hybridization.

Divergence between Parental Species

It is highly relevant to have an estimate of divergence between the focal species in the absence of hybridization. The parameter F_{st} is often reported in studies of anthropogenic hybridization, but is rarely used to motivate the marker density deployed for estimates of individual admixture, typically because the same markers are used to determine both F_{st} and individual Q estimates. Simulations have clearly shown that species (or subspecies) with lower divergence will require more markers to accurately estimate admixture, because of shared polymorphisms between them, leading to fewer **diagnostic markers** [33]. Although it might not be practical to use markers to estimate F_{st}, and then determine how many markers are necessary to estimate individual admixture scores, an initial assessment of F_{st} will hint at how much power a system has to detect advanced backcrosses.

Historical Admixture

Many systems have a history of repeated secondary contact and hybridization. Documenting historical admixture using genomic resources can determine whether the observed introgression is due to recent, anthropogenic forces or to natural causes, which will change the conservation status of the situation [34,35]. There are techniques for detecting historical admixture. For example, the ABBA-BABA test can be used to determine if there has been historical introgression, from a third species or population into each of two closely related sister taxa, to explain variation that is not well explained by a null assumption of bifurcating phylogeny [36]. This technique can be applied to either sequences of single individuals from each population, or to multiple individuals from each population [37], and can be used to indicate historical (hundreds to thousands of generations before present) admixture. Similarly, δaδi (diffusion approximation for demographic inference) analyses can be used to determine how well different demographic models fit the pattern of variation in the data, where demographic models can include admixture at different timepoints [38]. For example, demographic modeling was used to demonstrate that hybridization between golden-winged (*Vermivora chrysoptera*) and blue-winged warblers (*V. cyanoptera*) has probably been occurring since the original species split, and is not solely due to anthropogenic forces [39]. Finally, researchers can examine the length of haplotype blocks that are identical by descent because linkage disequilibrium decays over time as a result of recombination [40,41]. The distribution of haplotype block lengths should follow a Poisson distribution [41], and deviation from this distribution can be used to infer population admixture over both short (tens of generations) [42] and long timespans [41]. These and other techniques for disentangling historical and contemporary admixture are reviewed in [43].

Generations since Secondary Contact and Recombination Rates

It is important to estimate the number of generations since secondary contact to estimate the potential number of backcross generations in a system. This estimate might have substantial uncertainty, but in many cases of anthropogenic hybridization there are historical records suggesting when a non-native species was first introduced or sighted, and these can be combined with typical generation times for the taxa involved. The expected proportion of invasive genome in a backcrossed individual reduces by half with each successive generation of backcrossing [44].

Recombination taking place each generation leads to less linkage disequilibrium between non-native loci, which means that genotype at a species-specific marker in one position is less informative about surrounding, unsampled loci. For example, genomic regions with high recombination rates were found to be associated with more introgression of the non-native genome in replicate swordtail (*Xiphophorus birchmanni* and *X. malinche*) hybrid zones [17].

Owing to obligatory crossing over, which is expected to occur once per chromosome arm [45], at least twice as many markers as there are chromosome arms are necessary to cover each independent section of the genome. In some cases there is a species-specific estimate of recombination (e.g., [46]), or one can refer to taxon-specific patterns. For example, there is as much as 10-fold more recombination in avian genomes than in mammalian genomes [47]. In addition, information on recombination rate can be combined with genomic methods examining haplotype block lengths to allow dating of introgression events (as discussed above). We discuss how many markers are needed further in Box 1.

Assessing the Power of Markers

Many studies of anthropogenic hybridization assess the power of the genetic markers used by simulating hybrid genotypes, and then determining the power the markers have to detect these hybrid genotypes [48]. When assessing the power of markers in this way, it is important to ensure that the biology of the system is reflected in the simulation. In particular, if the two species of interest have been in contact for many generations, and F1s are thought to be fertile (Figure 1), then simulations should account for the possibility of many generations of backcrossing. This is rarely done in conservation genetic studies – many studies simulate backcrosses to assess the power of their markers, and find low power to detect even first-generation backcrosses, for example finding

Box 1. How Many Markers Do I Need To Discover Backcrossed Individuals in My System?

Substantial power is necessary to detect individuals that are the result of repeated generations of backcrossing. General rules have been suggested, including that twice as many markers are needed for each additional generation [44], and that at least 48 markers would be needed to consistently detect first-generation backcrossing in hybrids with parental species that have an F_{st} of 0.21 [33]. However, we are now in the age of genomics, when the cost of increasing marker density is dramatically decreasing [58], and thus marker numbers should be less of a barrier than previously. How many markers does a study then need to reliably detect backcrossed individuals?

To maximize detection of backcrossed individuals, researchers can increase their power in three ways: through increased divergence, the use of diagnostic markers, or increased numbers of markers. Studies with high divergence between hybridizing species have high power [33]. However, because many conservation biologists choose their study system based on conservation concerns and not to maximize power, this advice is not helpful. Diagnostic markers have fixed allelic differences between parental species and are the most powerful for backcross detection [25]. Ancestry-informative markers, those with strong allele-frequency divergence between species, are also very powerful [63]. Loci with weak allele-frequency divergence between species are least useful. Diagnostic and ancestry-informative markers can be determined based on genotyping and contrasting known parental species individuals, although this is not always feasible (e.g., [55]). In addition, the diagnostic properties of markers are a function of the populations and individuals that have been sampled; more extensive sampling sometimes demonstrates that selected markers are not diagnostic for all populations [66]. Generally speaking, the more markers that are used, the higher the likelihood of detecting admixture in an individual [33,44].

Assuming diagnostic markers, it is ideal to know the number of elapsed generations since the initial hybridization because, for every further generation of backcrossing, the proportion of introgressed genome reduces by half [44]. The number of generations since hybridization should be interpreted with an eye to policy. After some number of generations of unidirectional backcrossing, policy will dictate that we consider an individual to be parental species (again) [67]. It is best to make this decision before marker selection because it is impossible to apply policy decisions regarding the acceptability of backcrossed individuals if sufficient detection power is lacking.

If we are interested in all generations of backcrossing, then we can extend the deterministic model developed by Boecklen and Howard ([44]; equation 2) for the genomics era. We have made the same assumptions, specifically that backcrossing is unidirectional, loci are independent and Mendelian, all markers are diagnostic, all backcrossing is between the previous generation of backcrosses and parental species, and all genotypes are equally fecund [44]. We asked what proportion of backcrossed individuals are undetectable because they are homozygous for all diagnostic markers. We modeled 10 generations of backcrossing, and each of 10, 100, and 1000 diagnostic markers (Figure I). When using 10 diagnostic markers, 52% of fourth-generation backcrosses are homozygous for one parental species at all loci, and are thus undetectable as backcrosses. By contrast, 1000 diagnostic markers will allow powerful (85%) detection of ninth-generation backcrosses.

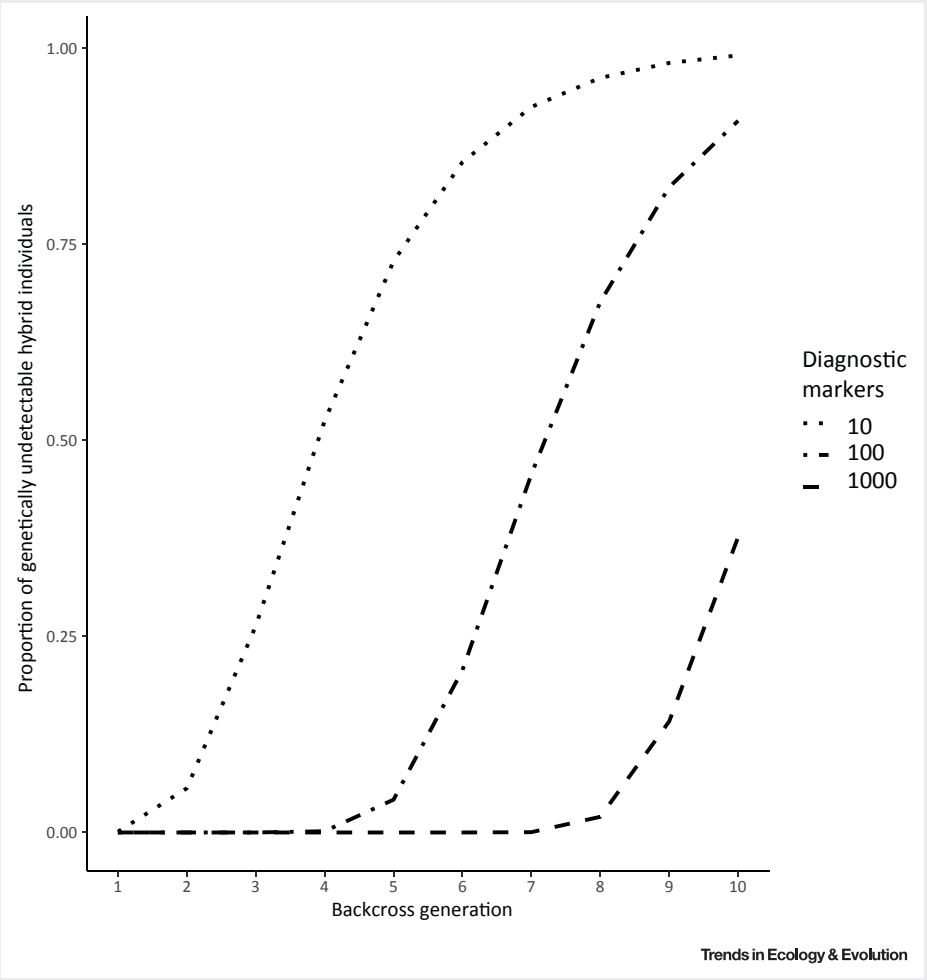

Figure I. An Extension of the Deterministic Model Presented by Boecklen and Howard [44]. The proportion of hybrid individuals that are homozygous at all the (diagnostic) markers, and are hence indistinguishable from the parental species that is being introgressed, increases with each generation of backcrossing, but decreases with increased marker density. This demonstrates that more markers than are typically used in studies of anthropogenic hybridization will be necessary to detect advanced backcrosses.

that <80% of first-generation backcrosses are properly assigned [49,50]. Further information obtained from laboratory or field studies, such as asymmetry in hybrid fertility (e.g., between sexes, Haldane's rule [51]; or according to the species of the mother of the F1, Darwin's corollary [52]), should also be included in simulations. For example, if previous laboratory work has established that backcrossing is largely unidirectional because of decreased fitness of hybrid individuals in the opposite direction (as expected by Darwin's corollary), or because of the relative abundances of the parental species, then mitochondrial markers should be integrated into future analyses to add power to detect hybrids.

Defining Hybrid Individuals

To be defined as a hybrid, a focal individual must be genetically differentiated from both parental species. Parental species are assumed to have an admixture (*Q*) score of 0 or 1, although, because of error (e.g., non-diagnostic markers, genotyping errors), very few individuals will

have an estimated score of exactly 0 or 1. Any score in between indicates a hybrid [25]. It is typical for a researcher to set a *Q* score as a cut-off for hybrid individuals, such that any individual above (or below) this score is considered parental. Thresholds are determined either by power – specifically, at what level can the markers differentiate between hybrids and parental species – or by the number of acceptably mismatched markers, for example one allele indicative of the other species might be an error, but two markers suggest hybridization [53]. These thresholds can range widely between studies, from 0.8 [54] to 0.999 [30] in relation to a parental species score of 1.0. Determination of the threshold is a balancing act between type I and type II errors, and the researcher must decide whether it is better to mistakenly assign a parental species individual as a hybrid (type I, 'accuracy' is too low [33]) or assign hybrid individuals as parental types (type II, 'efficiency' is too low [33]). If the researcher accepts a higher level of type II errors, they consider advanced backcrosses as parental species. For example, an admixture score threshold of 0.8 would include most second-generation backcrosses (87% of the genome is species A, 13% of the genome is species B, on average) as parental species. Similarly, with a *Q* score of 0.9, third-generation backcrosses (average of 93% species A) would be included as individuals of parental species.

There are two ways to diminish error introduced by using thresholds in species assignment. One obvious way is to employ more markers (Box 1), which increases the power of a study and allows the setting of thresholds approaching 0 and 1. Studies that have used thousands of markers use the most stringent thresholds (e.g., [30]). A second solution to the threshold problem is to eliminate thresholds entirely. Instead of assigning individuals to species classes based on point estimates, it is more appropriate to use **credible intervals** (or confidence intervals) around point estimates which capture uncertainty in the marker system appropriately (Box 2). In this scenario any individual with a credible interval overlapping 0 or 1 is considered to be a parental species, and all others are considered hybrids.

Box 2. Reporting Error

Credible (or confidence) intervals (CIs) are a powerful and intuitive way to assess confidence in the estimates being presented [68,69]. Measures of uncertainty are not always presented in estimates of anthropogenic hybridization (although see [53,70–73] for exceptions), perhaps because the uncertainty is very high when estimated. Credible intervals can be calculated using STRUCTURE [25], and standard errors can be calculated using ADMIXTURE [29]; reporting of error estimates is therefore easily implemented in a routine workflow.

There are practical implications of the reporting of credible intervals, particularly for individuals with very low or very high admixture values (*Q*). Cut-off thresholds have been used to determine if individuals are members of the parental populations or are admixed, but these thresholds are usually based on the detection power of a study (see main text). Because these are hard cut-offs, individuals with very similar levels of admixture can be assigned to different populations. For example, with a *Q* cut-off of 0.80, if individual 1 is assessed as $Q = 0.79$, it is determined to be admixed and, depending on the management of the system, may be culled. By contrast, if individual 2 is estimated to have $Q = 0.81$, it would be considered a parental species individual and be retained for breeding. There may be no substantive difference between these individuals, although this is impossible to tell using only point estimates.

We recommend that credible intervals should also be included in visual depictions of admixture. Typically, the key figure from a paper on anthropogenic hybridization is the characteristic 'STRUCTURE bar plot' [25] that uses stacked colors to denote genetic contributions from different source populations. These plots show the point estimates for each individual, and allow the author to determine thresholds for inclusion in each group. While such figures are compelling and easily interpreted, they do not convey the uncertainty around individual point estimates.

Allendorf and colleagues [11] noted that it is very difficult to make policy decisions when comparing different low point estimates of admixture. We recommend that researchers should focus on the uncertainty around *Q* estimates when making decisions about the genetic group each individual belongs to. It has been pointed out that the use of credible intervals demonstrates the high levels of uncertainty researchers are facing [70]. As they should! This problem will of course be substantially alleviated by using more markers (Box 1).

An additional problem in separating hybrid individuals from parental species is that some hybrids, particularly in later generations of backcrosses, will be homozygous for all sampled diagnostic loci by chance. This is due to increased variation in the proportion of genome inherited from each parental species with each generation of backcrossing [44] (Box 1). The hybrid nature of these individuals will be undetectable, and they will be classified as parental species, even though unmarked genome regions may be introgressed. Increasing the number of markers increases the probability of sampling a hybrid individual at loci that are heterozygous or homozygous for alleles representative of both parental species (Box 1).

Higher-Density Markers To Identify Bimodal Hybrid Zones

When researchers apply higher-density marker panels to examples of anthropogenic hybridization, they generally uncover more backcrossed individuals compared to studies using low-density panels, and can draw more accurate conclusions about the system. These newly detected backcrosses are often genetically very similar to the parental species, with <10% introgression, indicative of a bimodal hybrid zone. For example, in a study of Italian wolves (*Canis lupus italicus*) that hybridize with domestic dogs (*C. familiaris*), the use of 170 000 SNPs found that hybridization had occurred 3–5 generations before sampling [30]. This multigeneration backcrossing was not detectable in the population when 18 microsatellite markers were used [49]. Further, although very few individuals were found to have *Q* scores between 0.25 and 0.75, as would be expected in a hybrid swarm with a complete breakdown of reproductive isolation, 62% of sampled Eurasian wolves (*C. lupus lupus*) had a small proportion (<5%) of admixture with domestic dogs [55]. The Eurasian wolf–domestic dog system has a distribution of admixture scores and phenotypes that characterizes a bimodal hybrid zone with some degree of mating preference for parental phenotypes, or rare intermediate hybrids. In this system, most individuals are either phenotypically dog-like, with extreme *Q* scores at one end of the distribution, or phenotypically wolf-like, with *Q* scores at the other end of the distribution. There are few individuals with intermediate *Q* scores and phenotypes. This can be contrasted with the westslope cutthroat (*Oncorhynchus clarki lewisi*)–rainbow trout (*O. mykiss*) system, which has also recently been genotyped using 3180 diagnostic SNPs [56]. Although the increase in the number of markers did lead to increased detection of advanced backcrosses, there were also many individuals with intermediate *Q* scores and phenotypes [56,57]. This suggests that the westslope cutthroat–rainbow trout system is a hybrid swarm that has little assortative mating.

Designing an Ideal Study of an Anthropogenic Hybrid Zone

When embarking on a study of anthropogenic hybridization, there are many considerations in deciding on the genetic resources to be used (Box 1). As whole-genome sequencing (WGS) becomes cheaper [58], conservation biologists should consider whether WGS is the best way forward. First, WGS data allow the detection of heterogeneity of introgression across the genome. If conservation biologists truly adopt a 'gene viewpoint' of hybridization [16], then individuals ought to be classified based on whether they carry specific alleles at identified loci rather than on their overall *Q* scores (but see [10] for a discussion of the difficulty of implementing this approach). Second, WGS enables the researcher to distinguish between historical and contemporary introgression. Finally, we anticipate that the use of WGS will result in more diagnostic or **ancestry-informative markers** being detected, and thus make studies more powerful. Researchers will be more confident in their estimates of individual admixture, and will report the power and confidence associated with their analyses (Box 2). Although the bioinformatics skills required to assemble a genome and call SNPs may seem intimidating, we believe that (i) these are skills are now routinely taught in universities, and (ii) WGS presents an additional opportunity for conservation biologists to collaborate with

Box 3. Lessons from Natural Systems

Naturally occurring hybrid zones have long been used as 'natural laboratories' to study the speciation process [74]. The field of speciation genomics aims to understand how genomic differences build up to cause eventual reproductive isolation [75–78]. Recently, population geneticists have used genome-wide markers to ask questions regarding the genomic architecture of reproductive isolation and speciation, and how the genomes of diverged populations change in the face of ongoing gene flow [43,78,79]. Further, many studies of natural hybrid zones have focused on isolating signals from historical versus contemporary hybridization (main text and [78]). These questions that speciation biologists ask using hybrid zones could equally be asked in anthropogenic hybrid zones, particularly in studies that used WGS. Indeed, studies of anthropogenic hybrid zones may even have more power than those with naturally occurring secondary contact because, in some cases of introduced or escaped heterospecifics, phenotypic divergence is more extreme, meaning that fewer individuals would need to be sampled to enable, for example, admixture mapping [78].

Use of genomic data allows speciation geneticists to examine heterogeneity in divergence across the genome. Indeed, the questions we noted above are most interesting when heterogeneity is found. Genome scans look for regions of high divergence between species (F_{st} or d_{xy}) which may indicate regions that resist introgression, also known as 'speciation islands' [80] or 'islands of differentiation' [79]. Although such signals are not without controversy [81], and in some cases may represent phylogenetically derived regions of low recombination rather than reproductive isolation [82], they represent interesting candidate regions for fixed differences between hybridizing species, and thus could be used diagnostically by conservation biologists. For example, golden-winged (*Vermivora chrysoptera*) and blue-winged warblers (*V. cyanoptera*), which hybridize in eastern North America, are phenotypically distinct but undistinguishable when using low-density microsatellite marker panels [83]. Only with the use of WGS were six small divergent regions of the genome discovered, four of which are associated with either pigmentation or feather development genes and explain >90% of the variation in plumage [39]. This demonstrates that a focus on the use of high-density markers to explore heterogeneity across the genome allows higher power to both distinguish genetically between closely related, hybridizing species, and to associate genomic regions with diverged phenotypes, two possible goals of conservation biologists working on anthropogenic hybrid zones. We echo the call of [1] that conservation biologists can take a cue from speciation biologists that have, in many cases, developed methods that use genomics to ask interesting questions of hybrid zones.

Outstanding Questions

Do replicate anthropogenic hybrid zones show similar patterns of introgression? There are big evolutionary questions that could be answered by the types of data that conservation biologists working on anthropogenic hybridization could answer. For example, there are multiple replicate hybrid zones occurring in the wolf/dog, wild cat/domestic cat, red deer/sika deer, westslope cutthroat trout/rainbow trout systems. Even so, in many cases there is limited communication and collaboration between researchers, or different markers are used across studies [60]. Clearly this problem is not unique to this field, but it is the case that collaboration between researchers would be made easier with standardized genome-wide data aligned to a common genome. Genomic data make cross-study comparisons easier, and would allow easier comparison between studies.

Once there has been a breakdown of reproductive isolation characterized as hybridization, how common is maintenance of assortative mating within parental species? Is the strength of assortative mating stronger when species are more diverged, or perhaps between closely related species that have recently evolved reproductive isolation?

What is the relative frequency of hybrid swarms versus bimodal hybrid zones? We expect that the prevalence of bimodal hybrid zones has been underestimated because of the difficulty of detecting highly introgressed backcrosses. Increased use of high-density markers will make these cases easier to detect, and would enhance our understanding of systems that are bimodal hybrid zones.

speciation geneticists (Box 3). Another consideration is that high-quality DNA is needed for the most accurate assemblies, although progress is being made towards high-quality sequences from poor-quality samples (e.g., [59]). Although the use of WGS is more expensive than microsatellite marker studies, when the cost of microsatellite markers, including the cost of labor, was compared to the use of SNP markers in European wolves, SNPs were less expensive if at least 24 samples were genotyped [60]. This suggests that the use of thousands of variable genome-wide markers (e.g., from ddRAD [61]) may represent a practical middle ground for conservation biologists, depending on the history and biology of the system. Taken together, we believe that the best way forward to accurately detect backcrossing in studies of anthropogenic hybrid zones is to routinely use higher-density markers, including WGS when possible.

Concluding Remarks

Advanced backcrosses are unlikely to have been detected with many of the methods that biologists studying anthropogenic hybridization have used to date. Most studies of anthropogenic hybridization have used fewer than 20 markers [13], too few to reliably detect individuals that are backcrossed by more than two generations [33], unless the markers are perfectly species-diagnostic [44]. For this reason, it is rare for studies to consider backcrossed individuals past the second generation of backcrossing, regardless of the number of generations that have passed since secondary contact. We suggest here that studies should attempt to go much further. By accounting for the number of generations since secondary contact, and by increasing the density of genetic markers accordingly, many more backcrossed individuals will become distinguishable from the parental populations. We echo the call for more genetic markers to be used in these studies to allow higher accuracy and efficiency [1,3,13,33,62], particularly because we have now entered the genomics era, making tens or hundreds of

thousands of markers obtainable even in non-model systems [58]. It seems likely that anthropogenic hybridization will only increase in frequency and result in increased gene flow between previously isolated species [1]. An increase in the number of markers and associated power will also open up the opportunity to ask new questions in these systems, paralleling the questions speciation biologists explore in natural hybrid zones (Box 3). There are new challenges with increased marker density (see Outstanding Questions), but a genomic approach to studying these systems will help researchers to detect backcrosses and make the best policy recommendations.

Acknowledgments

We would like to thank Konrad Lohse and two anonymous reviewers for valuable feedback on the manuscript. S.E.M. was funded by the Swedish Research Council (Vetenskapsrådet, International Postdoc 2017-00499).

References

1. Grabenstein, K.C. and Taylor, S.A. (2018) Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends Ecol. Evol.* 33, 198–212
2. Brennan, A.C. *et al.* (2015) Hybridization due to changing species distributions: adding problems or solutions to conservation of biodiversity during global change? *Evol. Ecol. Res.* 16, 475–491
3. Wayne, R.K. and Shaffer, H.B. (2016) Hybridization and endangered species protection in the molecular era. *Mol. Ecol.* 25, 2680–2689
4. Crispo, E. *et al.* (2011) Broken barriers: human-induced changes to gene flow and introgression in animals. *Bioessays* 33, 508–518
5. Pimm, S.L. *et al.* (2006) The genetic rescue of the Florida panther. *Anim. Conserv.* 9, 115–122
6. Frankham, R. (2015) Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Mol. Ecol.* 24, 2610–2618
7. Suarez-Gonzalez, A. *et al.* (2016) Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. átrichocarpa* (black cottonwood). *Mol. Ecol.* 25, 2427–2442
8. Hamilton, J.A. and Miller, J.M. (2016) Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv. Biol.* 30, 33–41
9. Kovach, R.P. *et al.* (2016) Risk and efficacy of human-enabled interspecific hybridization for climate-change adaptation: response to Hamilton and Miller (2016). *Conserv. Biol.* 30, 428–430
10. Kardos, M. and Shafer, A.B. (2018) The peril of gene-targeted conservation. *Trends Ecol. Evol.* 33, 827–839
11. Allendorf, F.W. *et al.* (2001) The problems with hybrids: setting conservation guidelines. *Trends Ecol. Evol.* 16, 613–622
12. Allendorf, F.W. and Luikart, G. (2009) *Conservation and the genetics of populations,* John Wiley & Sons
13. Todesco, M. *et al.* (2016) Hybridization and extinction. *Evol. Appl.* 9, 892–908
14. Rhymer, J.M. and Simberloff, D. (1996) Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* 83–109
15. Rieseberg, L.H. *et al.* (1989) Hybridization in the island endemic, *Catalina mahogany*. *Conserv. Biol.* 3, 52–58
16. Petit, R.J. (2004) Biological invasions at the gene level. *Divers. Distrib.* 10, 159–165
17. Schumer, M. *et al.* (2018) Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 360, 656–660
18. Fitzpatrick, B.M. *et al.* (2010) Rapid spread of invasive genes into a threatened native species. *Proc. Natl. Acad. Sci.* 107, 3606–3610
19. Amish, S.J. *et al.* (2012) RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays. *Mol. Ecol. Resources* 12, 653–660
20. Epifanio, J. and Philipp, D. (2000) Simulating the extinction of parental lineages from introgressive hybridization: the effects of fitness, initial proportions of parental taxa, and mate choice. *Rev. Fish Biol. Fish.* 10, 339–354
21. Lamichhaney, S. *et al.* (2018) Rapid hybrid speciation in Darwin's finches. *Science* 359, 224–228
22. Jiggins, C.D. and Mallet, J. (2000) Bimodal hybrid zones and speciation. *Trends Ecol. Evol.* 15, 250–255
23. Rundle, H.D. and Nosil, P. (2005) Ecological speciation. *Ecol. Lett.* 8, 336–352
24. Allen, B. *et al.* (2016) Role of genetic background in the introgressive hybridization of rainbow trout (*Oncorhynchus mykiss*) with Westslope cutthroat trout (*O. clarkii lewisi*). *Conserv. Genet.* 17, 521–531
25. Pritchard, J.K. *et al.* (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959
26. Raj, A. *et al.* (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589
27. Kwan, Y.S. *et al.* (2014) Genomic replacement of native *Cobitis lutheri* with introduced *C. tetralineata* through a hybrid swarm following the artificial connection of river systems. *Ecol. Evol.* 4, 1451–1465
28. Le Roux, J.J. *et al.* (2015) Genetic analysis shows low levels of hybridization between African wildcats (*Felis silvestris lybica*) and domestic cats (*F.* s. catus) in South Africa. *Ecol. Evol.* 5, 288–299
29. Alexander, D.H. and Lange, K. (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* 12, 246
30. Galaverni, M. *et al.* (2017) Disentangling timing of admixture, patterns of introgression and phenotypic indicators in a hybridizing wolf population. *Mol. Biol. Evol.* 34, 2324–2339
31. Nielsen, E.E. *et al.* (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Mol. Ecol. Resour.* 6, 971–973
32. van Heugten, R.A. *et al.* (2017) Sleeping with the 'enemy': hybridization of an endangered tree weta. *Conserv. Genet.* 18, 1377–1387
33. Vähä, J.-P. and Primmer, C.R. (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Mol. Ecol.* 15, 63–72
34. vonHoldt, B.M. *et al.* (2016) Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Sci. Adv.* 2, e1501714
35. Hohenlohe, P.A. *et al.* (2017) Comment on 'Whole-genome sequence analysis shows two endemic species of North American wolf are admixtures of the coyote and gray wolf'. *Sci. Adv.* 3, e1602250

36. Durand, E.Y. *et al.* (2011) Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252
37. Soraggi, S. *et al.* (2018) Powerful inference with the D-statistic on low-coverage whole-genome data Genes, Genomes. *Genetics* 8, 551–566
38. Gutenkunst, R.N. *et al.* (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5, e1000695
39. Toews, D.P. *et al.* (2016) Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr. Biol.* 26, 2313–2318
40. Gravel, S. (2012) Population genetics models of local ancestry. *Genetics* 191, 607–619
41. Harris, K. and Nielsen, R. (2013) Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet.* 9, e1003521
42. Palamara, P.F. *et al.* (2012) Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* 91, 809–822
43. Payseur, B.A. and Rieseberg, L.H. (2016) A genomic perspective on hybridization and speciation. *Mol. Ecol.* 25, 2337–2360
44. Boecklen, W.J. and Howard, D.J. (1997) Genetic analysis of hybrid zones: numbers of markers and power of resolution. *Ecology* 78, 2611–2616
45. Dumont, B.L. (2017) Variation and evolution of the meiotic requirement for crossing over in mammals. *Genetics* 205, 155–168
46. Johnston, S.E. *et al.* (2018) A genomic region containing REC8 and RNF212B is associated with individual recombination rate variation in a wild population of red deer (*Cervus elaphus*). *G3* 8, 2265–2276
47. Ellegren, H. (2010) Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* 25, 283–291
48. Anderson, E.C. *et al.* (2009) Statistical methods for identifying hybrids and groups. In *Population Genetics and Animal Conservation* (Bertorelle, G., ed.), pp. 25–41, Cambridge University Press
49. Randi, E. (2008) Detecting hybridization between wild species and their domesticated relatives. *Mol. Ecol.* 17, 285–293
50. Galov, A. *et al.* (2015) First evidence of hybridization between golden jackal (*Canis aureus*) and domestic dog (*Canis familiaris*) as revealed by genetic markers. *R. Soc. Open Sci.* 2, 150450
51. Haldane, J.B. (1922) Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* 12, 101–109
52. Turelli, M. and Moyle, L.C. (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176, 1059–1088
53. Senn, H.V. and Pemberton, J.M. (2009) Variable extent of hybridization between invasive sika (*Cervus nippon*) and native red deer (*C. elaphus*) in a small geographical area. *Mol. Ecol.* 18, 862–876
54. Schulte, U. *et al.* (2012) Rapid genetic assimilation of native wall lizard populations (*Podarcis muralis*) through extensive hybridization with introduced lineages. *Mol. Ecol.* 21, 4313–4326
55. Pilot, M. *et al.* (2018) Widespread, long-term admixture between grey wolves and domestic dogs across Eurasia and its implications for the conservation status of hybrids. *Evol. Appl.* 11, 662–680
56. Hohenlohe, P.A. *et al.* (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Mol. Ecol.* 22, 3002–3013
57. Boyer, M.C. *et al.* (2008) Rainbow trout (*Oncorhynchus mykiss*) invasion and the spread of hybridization with native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). *Can. J. Fish. Aquat. Sci.* 65, 658–669
58. Ellegren, H. (2014) Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* 29, 51–63
59. van der Valk, T. *et al.* (2017) Whole mitochondrial genome capture from faecal samples and museum-preserved specimens. *Mol. Ecol. Resour.* 17, e111–e121
60. de Groot, G.A. *et al.* (2016) Decades of population genetic research reveal the need for harmonization of molecular markers: the grey wolf *Canis lupus* as a case study. *Mamm. Rev.* 46, 44–59
61. Andrews, K.R. *et al.* (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17, 81–92
62. Koskinen, M.T. *et al.* (2004) The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. *Hereditas* 141, 61–67
63. Shriver, M.D. *et al.* (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum. Genet.* 112, 387–399
64. Wright, S. (1943) Isolation by distance. *Genetics* 28, 114–138
65. Avise, J.C. (2012) *Molecular Markers, Natural History and Evolution,* Springer Science & Business Media
66. Hand, B.K. *et al.* (2015) Genomics and introgression: discovery and mapping of thousands of species-diagnostic SNPs using RAD sequencing. *Curr. Zool.* 61, 146–154
67. Lavretsky, P. *et al.* (2016) Becoming pure: identifying generational classes of admixed individuals within lesser and greater scaup populations. *Mol. Ecol.* 25, 661–674
68. Altman, D. *et al.* (2013) *Statistics with confidence: confidence intervals and statistical guidelines,* John Wiley & Sons
69. Nakagawa, S. and Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev.* 82, 591–605
70. Bohling, J.H. and Waits, L.P. (2011) Assessing the prevalence of hybridization between sympatric *Canis* species surrounding the red wolf (*Canis rufus*) recovery area in North Carolina. *Mol. Ecol.* 20, 2142–2156
71. Trigo, T. *et al.* (2008) Inter-species hybridization among neotropical cats of the genus *Leopardus*, and evidence for an introgressive hybrid zone between *L.* geoffroyi and *L. tigrinus* in southern Brazil. *Mol. Ecol.* 17, 4317–4333
72. Beaumont, M. *et al.* (2001) Genetic diversity and introgression in the Scottish wildcat. *Mol. Ecol.* 10, 319–336
73. Yokoyama, R. *et al.* (2009) Disturbance of the indigenous gene pool of the threatened brook lamprey *Lethenteron* sp. S by intraspecific introgression and habitat fragmentation. *Conserv. Genet.* 10, 29–43
74. Hewitt, G.M. (1988) Hybrid zones – natural laboratories for evolutionary studies. *Trends Ecol. Evol.* 3, 158–167
75. Nosil, P. and Feder, J. (2012) Genomic divergence during speciation, causes and consequences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 332–342
76. Seehausen, O. *et al.* (2014) Genomics and the origin of species. *Nat. Rev. Genet.* 15, 176
77. Butlin, R.K. (2010) Population genomics and speciation. *Genetica* 138, 409–418
78. Nadeau, N.J. and Kawakami, T. (2019) Population genomics of speciation and admixture. In *Population Genomics* (Rajora, O.P., ed.), pp. 1–41, Springer
79. Wolf, J.B. and Ellegren, H. (2017) Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* 18, 87
80. Turner, T.L. *et al.* (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3, e285
81. Cruickshank, T.E. and Hahn, M.W. (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23, 3133–3157
82. Burri, R. *et al.* (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* 25, 1656–1665
83. Vallender, R. *et al.* (2007) Complex hybridization dynamics between golden-winged and blue-winged warblers (*Vermivora chrysoptera* and *Vermivora pinus*) revealed by AFLP, microsatellite, intron and mtDNA markers. *Mol. Ecol.* 16, 2017–2029